Adsorption of the natural protein surfactant Rsn-2 onto liquid interfaces

Journal Name

ARTICLE TYPE

Cite this: DOI: 10.1039/xxxxxxxxxx

Adsorption of the Natural Protein Surfactant Rsn-2 onto Liquid Interfaces

Giovanni B. Brandani[a], Steven J. Vance[b], Marieke Schor[a], Alan Cooper[b], Malcolm W. Kennedy[c], Brian O. Smith[d], Cait E. MacPhee[a], and David L. Cheung[e,f]

Received Date
Accepted Date

DOI: 10.1039/xxxxxxxxxx

www.rsc.org/journalname

Abstract

In order to stabilise foams, droplets, and films at liquid interfaces nature has evolved a diverse range of protein biosurfactants. In contrast with synthetic surfactants, these protein surfactants combine surface activity with biocompatibility and low aggregation in solution. One recently studied example is the protein Rsn-2, a component of the foam nest of the tropical frog *Physalemus pustulosus*, which has been predicted to undergo a clamshell-like opening transition at the air-water interface. Using atomistic molecular dynamics simulations and surface tension measurements the adsorption of Rsn-2 onto air-water and cyclohexane-water interfaces is studied. The protein adsorbs readily at both interfaces, with the interfacial adsorption being mediated by the hydrophobic N-terminus. At the cyclohexane-water interface the clamshell opening is observed, due to the more favourable interaction between hydrophobic residues and cyclohexane molecules, along with penetration of cyclohexane molecules into the protein core. Simulations of proteins with the N-terminus deleted showed that this can inhibit adsorption onto the air-water interface. This is consistent with experimental measurements of the surface tension, which found that N-terminal deletions lead to slower decrease in the surface tension and higher limiting surface tension values. Curiously, deletion of the hydrophilic C-terminus can also affect adsorption, suggesting that this plays a role in orienting the protein in order to optimise adsorption. The detailed characterisation of the interfacial behaviour gives insight into the factors that control interfacial adsorption of proteins. We anticipate that this may inform new applications of this and similar proteins in areas including drug delivery and food technology and may also be used in the design of synthetic molecules showing similar changes in conformation at interfaces.

1 Introduction

Due to the ubiquity of liquid interfaces in nature and their importance in a number of biological processes, a variety of proteins have evolved to function in interfacial environments. Among these are protein biosurfactants, surface active proteins that act to reduce surface tension[1–3], and have the ability to stabilise foams and emulsions. Alongside their surfactant properties these have typically evolved to be biocompatible which has led them to exhibit structures that are quite different to the standard polar head/non-polar tail of most synthetic surfactants[4,5]. Possibly the best-known and studied example of biosurfactant proteins are the hydrophobins[3], small, amphiphilic proteins expressed by filamentous fungi (along with similar bacterial proteins[6]). These are characterised by a hydrophobic patch on one face, giving them an amphiphilic structure similar to surfactants or Janus particles[7]. Caseins are a class of micelle forming proteins[8], that are an example of intrinsically disordered proteins. These contain regions of hydrophobic and hydrophilic amino acids, having structures reminiscent of block copolymers. A final example are lung surfactant proteins[1], in particular surfactant proteins SP-B and SP-C. These are small α-helical proteins, with their surface activity arising due to the amphipathic nature of their helices, that play a role

[a] *School of Physics and Astronomy, University of Edinburgh, Edinburgh EH9 3FD, United Kingdom. Email: c.e.macphee@ed.ac.uk*
[b] *School of Chemistry, University of Glasgow, Glasgow, G12 8QQ, UK*
[c] *School of Life Sciences, University of Glasgow, Glasgow, G12 8QQ, UK*
[d] *Institute of Molecular, Cell, and Systems Biology, University of Glasgow, G12 8QQ, UK. Email: brian.smith@glasgow.ac.uk*
[e] *Department of Pure and Applied Chemistry, University of Strathclyde, Glasgow, G1 1XL, UK*
[f] *School of Chemistry, National University of Ireland Galway, Galway, Ireland. Email: david.cheung@nuigalway.ie*

in modulating the surface tension of pulmonary fluid.

An interesting example of biosurfactant proteins is Rsn-2[2], a protein found in the foam nests of the tropical frog *Physalemus pustulosus*. Rsn-2 was found to be the main surfactant component of this mixture with its removal inhibiting foam formation. Its sequence is unlike any other surfactant protein and it exhibits an amphiphilic character, with a hydrophobic region at the N-terminus (residue ids 1 to 9: LILDGDLLK) and a highly polar C-terminus (residue ids 87 to 96: RKDDDDDDGY). However, high-resolution NMR studies have shown that the amphiphilic character expected from the sequence and surface activity is not apparent in the solution structure. The solution structure of Rsn-2 (PDB ID: 2WGO) comprises two regular secondary structure features, consisting of an α-helix and a four-stranded β-sheet, joined by a flexible linker region and with flexible N- and C-terminal tails. Unlike the hydrophobins, no hydrophobic patches were found on the protein surface, which is consistent with the lack of oligomerisation in solution[9]. Instead, studies of layers of Rsn-2 at the air-water interface suggest a more sophisticated mechanism for its surface activity. Neutron reflectivity measurements found a layer thickness of $\sim$8-10 Å[2], thinner than would be expected from its solution structure. Combined with the flexible linker region seen in the solution structure, MacKenzie *et al*[2] suggested that the protein opens up at the interface separating the helix and sheet regions, allowing the hydrophobic amino acids in the protein core to become exposed to air or oil. Further support for this comes from polarised IRRAS measurements, which showed that the α-helix and β-sheet secondary structures remain intact and lie in the plane of the interface.

Moreover, both N- and C-terminal regions of the protein were shown to be more dynamic than the globular core and they may be able to participate in the interfacial adsorption. Indeed, previous experiments and simulations with a highly coarse-grained model[10] suggested that the N-terminal tail can interact with the interface before the protein core exposes its hydrophobic residues. While this is an appealing model for interfacial adsorption, the precise atomistic details for interfacial attachment and opening are still unclear, with the role of the N- and C-terminal regions being of particular interest.

In this paper atomistic molecular dynamics (MD) simulations and biophysical analysis of deletion mutants are used to investigate the adsorption of Rsn-2 onto air-water and cyclohexane-water interfaces. In recent years MD has been used to investigate the interfacial adsorption of a number of proteins, including several hydrophobins[7,11–13], whey proteins, such as β-lactoglobulin[14,15] and barley LTP[16], lysozyme[17], and myoglobin fragments[18], giving insight into their structures at interfaces and the relationship between protein structure and interfacial behaviour. Here we examine the mechanism of interfacial attachment of Rsn-2, and in the case of a water-cyclohexane interface, observed the hypothesised clamshell opening. Through consideration of mutant proteins with sections of the N- and C-termini deleted we investigate the role of these regions on interfacial attachment. Finally, we estimate the free energy contributions to the adsorption from the interactions between the protein and the two fluid phases and from the interfacial tension (see methods section for more details on these estimates).

2 Methods

2.1 Molecular dynamics simulations

All simulations were performed with the molecular dynamics software GROMACS 4[19]. At the air-water interface the protein was modelled using the CHARMM27 force field[20,21], whereas at the oil-water interface we used a version of the AMBER99sb-ildn force field[22] that includes a parametrisation for cyclohexane molecules[23]. To test the dependence of the results on the force field used, simulation of the wild type Rsn-2 at the air-water interface using the AMBER99sb-ildn force field was performed, giving results consistent with those from the CHARMM27 force field. In all cases water was represented according to the TIP3P model[24].

To prepare the system for the adsorption simulations, a protein configuration taken from the NMR structure is centred and solvated in a cubic box with side lengths set so that the protein is at least 1.2 nm away from the periodic image of the box (in oil-water simulations the size of the simulation box was further increased so that this condition is satisfied even upon unfolding). Counter ions were added to neutralise protein charge and the size of the box was increased along the z direction to create the two interfaces; in the oil-water system, the additional empty space is then filled with cyclohexane molecules. For each type of protein and interface considered we performed three production runs, with different random starting protein conformations from the NMR structure used for each (for the air-water interface we used frames 1, 11 and 21 for runs 1, 2 and 3 respectively, whereas for the oil-water interface we used frames 6, 14, and 16). The system is equilibrated in the following steps: a steepest-descent energy minimisation procedure, a 50 ps run in the NVT ensemble and, only for the oil-water interface, a 100 ps run in the NAP_zT ensemble; in the last two steps the heavy atoms of the protein are restrained at their initial value using harmonic springs with a force constant equal to 1000 kJ mol^{-1} Å^{-2}. Production runs were performed in the NVT ensemble for the air-water interface, and in the NAP_zT ensemble for the oil-water interface.

In all cases (including equilibration), the temperature was maintained at 300 K using the velocity-rescale thermostat[25], and in the oil-water system the pressure along the z direction was kept at 1 bar using the Parrinello-Rahman algorithm[26]. The equations of motion were integrated using a timestep of 2 fs. All bonds involving hydrogen atoms were constrained at their equilibrium position using the LINCS algorithm[27], whereas the geometry of water was constrained using the SETTLE algorithm[28]. Electrostatic interactions were treated with the Particle mesh Ewald algorithm[29] with a Fourier grid spacing of 1.5 Å and a short-range cut-off of 10 Å. For simulations with the Amber force field short-range Van der Waals interactions were switched to zero from 8 to 9 Å, while for Charmm the Van der Waals interactions are cut-off at 12 Å.

Secondary structure assignments were made using the STRIDE algorithm[30]. The partition free energy ($\Delta F_{\text{partition}}$) in Table 1 was estimated as the sum of the air-water or water-cyclohexane transfer free energies[31] of each residue for which the side-chain centre

of mass position is located inside the hydrophobic phase, using the condition $z > z_G$, where the position of the interface z_G was computed from the Gibbs dividing surface for the air-water interface[32] and from the point at which the density of water equals the density of cyclohexane for the oil-water interface. The residues that form the hydrophobic core of the protein were not included in the sum, since we do not expect a significant gain in energy when these residues are transferred from the already hydrophobic environment of the core to the apolar fluid phase. These were identified as those that are not part of the N-terminal tail (id 1 to 16, which are known to be flexible from the NMR structure) and exposing a surface area lower than 30 Å in all conformations of the NMR structure. To estimate the surface area occupied by the protein at the interface, a grid of points is placed on the interface, defined by z_G, with the area being estimated from the number of grids points that were within a cut-off distance of $R^p_{vdw} + R^O_{vdw}$ from any protein atom, where R^p_{vdw} is the van der Waals radius of the protein atom and $R^O_{vdw} = 1.4$ Å is the van der Waals radius of the water oxygen. All calculations presented in Table 1 used a grid spacing of 1 Å in the x and y directions (calculations with more grid points gave essentially identical results). The surface tension was taken from previous simulations[33] for the air-water interface, whereas for the water-cyclohexane interface it was computed directly from a molecular dynamics simulation in the absence of the protein using:

$$\gamma = 0.5 L_\perp \langle P_\perp - P_\parallel \rangle \quad , \tag{1}$$

where $L_\perp$ is the size of the simulation box along the direction perpendicular to the interface, $P_\perp$ and $P_\parallel$ represent respectively the components of the pressure perpendicular and parallel to the interface, and $\langle\rangle$ represents an average over time during an equilibrium simulation.

In Fig. 9, the number of contacts between the hydrophobic core and cyclohexane, and between α-helix and β-sheet are given by:

$$N_{\text{contacts}} = \sum_{ij} \frac{1-(r_{ij}/r_0)^6}{1-(r_{ij}/r_0)^{12}} \tag{2}$$

where r_{ij} are respectively the distances between the C_γ atoms of the residues in the core and the cyclohexane molecules, and the distance between the C_γ atoms in the α-helix and those in the β-sheet we only include the large hydrophobic residues of the core, which possess a C_γ atom. The sum runs over the appropriate pair of atoms, and $r_0 = 5$ Å. The switching function on the right goes to 0 at large distances and 1 at short distances, therefore identifying the formation of a hydrophobic contact.

For the clustering of protein orientations in Fig. 8, we employed the recent algorithm of search and find of density peaks[34]. In order to characterise the protein orientation, we considered the z-coordinates of residue ids 16 to 88 relative to the centre of mass of this region; N- and C-terminal tails have been excluded because of their flexibility. The distance between any pair of conformations was then defined as the root mean squared distance between the corresponding vectors of z coordinates just defined. The three clusters in Fig. 8 have been obtained using a cutoff distance of 11 Å , but the identification of the clusters is robust with respect to changes in this value.

2.2 Recombinant Rsn-2 production

Recombinant Rsn-2 was produced in *E. coli* and purified as previously described[2]. Briefly, BL21[DE3] cells transformed with a pET28 based plasmid encoding Rsn-2 as a fusion protein with a thrombin cleavable N-terminal His_6-tag were grown in Luria-Bertani medium and induced with IPTG. The protein was purified from the soluble fraction of the lysed cells by Ni^{2+}-affinity chromatography and the tag removed by digestion with thrombin followed by a further Ni^{2+}-affinity chromatography step and size exclusion chromatography on a Superdex75 column (GEHealthcare). N- and C-terminal deletion mutants were produced in a similar fashion. After thrombin cleavage, the 'wild type' and truncated Rsn-2 proteins have an additional N-terminal tetrapeptide sequence, GSHM, before their first residue of the native sequence.

2.3 Surface activity measurements

Surface tension measurements were performed using a Kibron Microtrough S (Kibron, Espoo, Finland). The instrument was calibrated against air and pure water as per the manufacturer's instructions and then zeroed against buffer. 500 μL samples were placed in a well of the multi-well plate immediately before measurement. The probe was immersed in the solution to ensure complete wetting, raised and then lowered until in contact with the surface before recording was initiated. The time taken from sample deposition to the start of recording was less than 10 s. Between samples the probe was flamed and between experiments, the multi-well plate was washed extensively with ethanol followed by ultrapure water and dried thoroughly.

Protein samples were created by serial dilutions into 20 mM sodium phosphate, 50 mM sodium chloride, 1 mM sodium azide, pH 7.5 buffer from a stock solution in the same buffer. Of the protein concentrations tested, two were chosen to illustrate the differences between the Rsn-2 variants. At 89 μM (equivalent to 1000 μg/mL for full length Rsn-2), surface activity is essentially fully developed within the dead time of the experiment allowing the ultimate effectiveness of the variants as surfactants to be assessed. In contrast, 445 nM is a near limiting concentration for the activity of wild type Rsn-2, with surface tension depression developing over several hundred seconds allowing the kinetics of its development to be compared for different variants.

3 Results

3.1 Wild-type Rsn-2

We first consider the behaviour of the protein at an air-water interface. Consistent with experimental observations, simulations show that Rsn-2 adsorbs readily onto the interface, with this occurring within 50 ns for all three replicate runs (Fig. 1). From the centre-of-mass positions it can be seen that the protein can contact the interface without becoming permanently attached. For example for the first run the protein first contacts the interface within approximately 5 ns (Fig. 1B) but remains at the interface for less than 1 ns before diffusing back into bulk solution. The protein makes another transient contact with the interface before becoming permanently attached from about 38 ns onwards (Fig. 1D). In the adsorbed configuration most hydrophobic side-

chains located in the N-terminal tail are exposed to the air phase. While the timescales differ for the other replicate runs similar behaviour is seen in all cases.

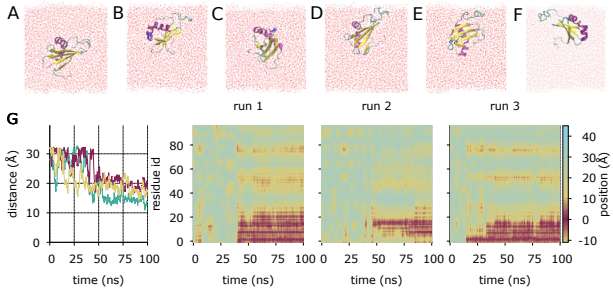

Fig. 1 Adsorption of wild type Rsn-2 at an air-water interface. (A-F) Snapshots of Rsn-2 adsorption observed in the first simulation run at times (left to right) 0, 5, 15, 38, 55 and 100 ns. (G) Distance of the protein centre of mass from the interface and residue-interface separations as a function of time. Cyan, red, and yellow lines denote first, second, and third runs respectively.

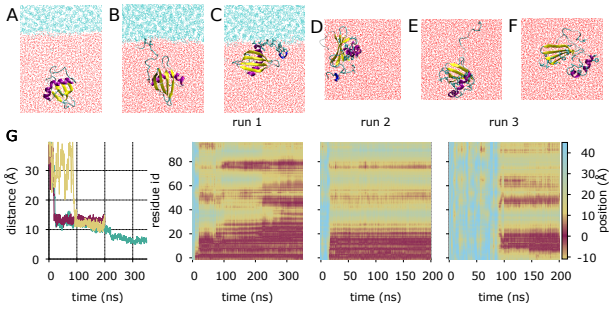

Fig. 2 Adsorption of wild type Rsn-2 at a cyclohexane-water interface. (A-F) Snapshots of Rsn-2 adsorption observed in the first simulation run corresponding, from left to right, to times 0, 12, 24 (side view and top view), 131 and 263 ns. (G) Distance of the protein centre of mass from the interface and residue-interface separations as a function of time. Cyan, red, and yellow lines denote first, second, and third runs respectively.

At a water-cyclohexane interface the initial stages of the adsorption of wild type Rsn-2 are qualitatively similar to those observed at an air-water interface. Permanent contacts are formed within 100 ns, and in all three runs the N-terminal tail is the part of the protein that contributes the most to the initial adsorption (Fig. 2). For instance in the first simulation run (Fig. 2B) at 12 ns the N-terminal tail interacts with the interface through residues L1 and I2, and from 24 ns (Fig. 2C) all hydrophobic residues in the tail (up to V16) have their side-chains exposed to the oil phase. The orientation of the protein at the interface can also change as a function of time and depends on the simulation run; for example from snapshots of the system observed from the direction perpendicular to the interface into the water phase, we can see that at 24 ns (Fig. 2D) both the helix and the β-sheet are perpendicular to the interface, whereas at 131 ns (Fig. 2E) they are parallel to the interface. This change in orientation corresponds to new contacts with the interface formed by the hydrophobic residues I17, L20 and F21 located at the beginning of the helix.

In this simulation run the protein also undergoes a large scale conformational change that is not observed in runs 2 and 3 or at the air-water interface. Starting from the formation of new contacts with residues L25, F29, V78 and P79, the cyclohexane molecules progressively invade the hydrophobic core of the protein and the distance between the β-sheet and the α-helix increases, as can be seen in the snapshot taken at 263 ns. This partial unfolding of the protein can be described as an unhinging of the helix with respect to the β-sheet, in which the core becomes directly exposed to the oil phase without making unfavourable contacts with water. The length of this simulation was extended up to 350 ns until the protein structure remained stable. This conformational change causes an increase in the area occupied by the protein at the interface, leading to a higher surface coverage for a given amount of protein. This is likely to be important for the biological role of Rsn-2 in foam stabilisation[9] and supports the mechanism suggested from previous experimental studies of Rsn-2 at an air-water interface[2].

The role of each part of the protein in the adsorption can be understood from the separation between the centre of mass of each residue and the interface as a function of time, as shown in Figs. 1G and 2G. For both air-water and oil-water interfaces it can be seen that the N-terminus is always close to the interface at the end of the simulation. The importance of the N-terminus can be explained by its high hydrophobicity (there are 6 leucines, 2 isoleucines, 1 proline and 1 valine within the first 17 residues of the protein) and flexibility. The latter is highlighted by the the high root mean square deviation with respect to the initial conformation (Fig. 4) and by the fact that the first 16 residues of the protein were determined to be disordered from NMR experiments[2]. Because of the flexibility of the tail, the side chains of these residues are often exposed to the solvent, and they can easily adsorb at the interface. This behaviour is similar to the fly-casting mechanism observed in intrinsically disordered proteins[35]. Other regions of the protein can also form contacts with the interface. For instance, at the oil-water interface, in the third simulation run the protein adopts a perpendicular orientation with respect to the interface, different to the orientation observed in the other runs. This orientation appears to be stabilised by a hydrophobic contact with the interface formed by residue L47, which does not appear in runs 1 and 2. The hydrophobic residues at the beginning of the α-helix (L20 and F21) interact with the interface in runs 1 and 2 at the oil-water interface and in runs 1 and 3 at the air-water interface, and they seem to be responsible for the parallel orientation of the protein. Residues V78 and P79 are also close to the interface in most simulation runs, and their interaction anticipates the unhinging of the core observed in the first run at the oil-water interface.

The observation of different stable orientations at the interface might explain why the unhinging is a rare event. This suggests that this process has to proceed via the crossing of one or more free-energy barriers, corresponding to adopting the optimal interaction with the interface that will enable the unhinging motion of the core. The contribution of hydrophobic residues outside the tail might also decide whether adsorption is permanent or only

transient. For instance in the first simulation run at the air-water interface the initial attachment (at ~ 5 ns) involves purely the N-terminal region (up to residue 18) whereas the later, permanent adsorption involves additional residues (K42-Y56 and V76-P79). As discussed above, for the water-cyclohexane simulations the attachment of these residues to the interface, as well as aiding permanent adsorption, may also precede the unhinging of the protein.

Shown in Fig. 3 is the secondary structure of the protein during attachment to the interface. This is largely unchanged, with both the long α-helix (residues 19 to 38) and the β-sheet (residues 45 to 88) remaining intact. The only partial loss of secondary structure is observed in the α-helix during the unhinging of the protein in the first run at the oil-water interface, and even in this case the conformation of this region is very close to the initial one. This maintenance of native structure is indicative of the fact that this protein has evolved to function at liquid interfaces so retains its structure here, in a similar manner to many other biosurfactant proteins such as the hydrophobins[3]. The conservation of secondary structure upon adsorption has also been observed from recent MD simulations of the hydrophobin HFBI[13], the whey protein β-lactoglobulin[14], and a peptide derived from myoglobin[18].

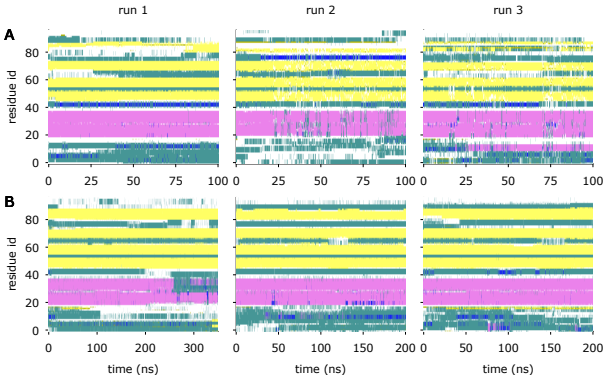

Fig. 3 Secondary structure per residue of wild-type Rsn-2 at (A) air-water and (B) oil-water interfaces as a function of time (magenta α-helix, blue 3/10 helix, yellow β-sheet, green turn).

The root mean square deviation (rmsd) of Rsn-2 (Fig. 4) shows the flexibilities of different parts of the protein, with the N-terminal tail being the most flexible. The structured core of the protein (residues 16 to 88) is typically very rigid, with the exception of the first run at the oil-water interface, where the large increase in rmsd illustrates the unhinging transition.

3.2 Effect of N and C-terminal deletions

To examine the contribution of the N and C-termini to the adsorption of Rsn-2 the surface activity of five mutants, ΔL1-L3, ΔL1-P15 (N-terminus), ΔY96 and ΔD89-Y96 (C-terminus), and one double mutant, ΔL1-P15+ΔD89-Y96, was measured by microtrough tensiometry. This was compared to that of wild type Rsn-2 at concentrations of 445 nM, to allow kinetic effects to be observed, and 8.9 μM, to assess the effect on overall surface activity (Fig. 5). Thus the specific contributions to surfactant activity of each of the

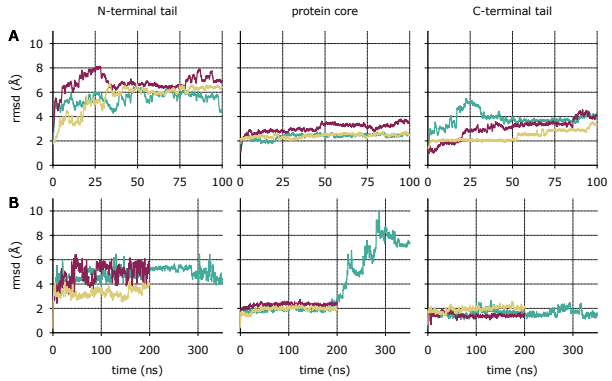

Fig. 4 Backbone root mean square displacement from the initial conformation of wild-type Rsn-2 at (A) air-water and (B) oil-water interfaces for the N-terminal tail (from the tag to residue id 15), the structured core (residues id 16-88) and the C-terminal tail (from residue id 89).

motifs within the termini could be dissected. N-terminal deletions ΔL1-L3 and ΔL1-P15 delay the kinetics of surface tension depression at a concentration of 445 nM and reduced the overall activity by 3-5 mN/m at 8.9 μM. The ΔL1-P15 deletion slows the kinetics of Rsn-2 activity considerably more than the ΔL1-L3 mutant and reduces the decrease in surface tension by about 2 mN/m at the higher concentration.

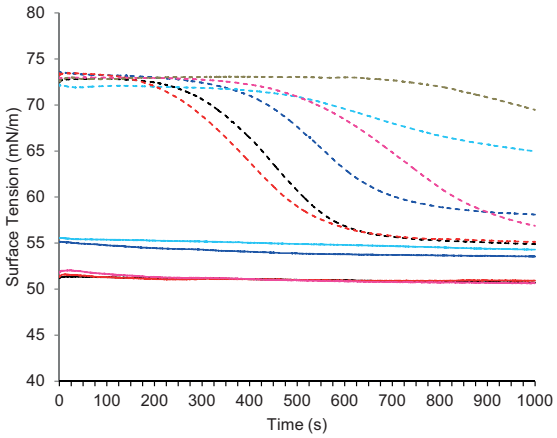

Fig. 5 The kinetics of development of surface tension depression by Rsn-2 and its N- and C-terminal deletion mutants measured by microtrough tensiometer. Representative curves for wild type (black), ΔL1-L3 (blue), ΔL1-P15 (cyan), ΔY96 (red), ΔD89-Y96 (pink) and ΔL1-P15+ΔD89-Y96 (brown) Rsn-2 at limiting (445 nM; dashed lines) and saturating (89 μM; solid lines) concentrations are shown.

Of the C-terminal deletions, the ΔY96 mutation does not appear to have any significant effect upon surface activity, and the ΔD89-Y96 reaches final surface tensions indistinguishable from WT Rsn-2 at both concentrations. However, at the lower concentration, ΔD89-Y96's kinetics lie between those of the two N-terminal mutants. The ΔL1-P15+ΔD89-Y96 double mutant displays the slowest kinetics of all the termini mutants, with activity only beginning to be evident after 600 seconds at the lower concentration. The minimum surface tension reached by this mutant (57.5 mN/m) is very similar to that of ΔL1-P15 alone, supporting

the inference that the C-terminal mutants affect the kinetics of development of, but not the final surfactant activity achieved.

We therefore tested how deletions of the N-terminal first 3 (ΔL1-L3) and first 15 (ΔL1-P15) residues affect interfacial attachment (Fig. 6). Although the highly charged C-terminus does not directly interact with the interface, it may play a role in correctly orienting the protein, thus facilitating attachment. To test this hypothesis, we also studied mutants of RSN-2 in which the last 1 (ΔY96) and last 7 (ΔD89-Y96) residues were deleted. Figure 6 shows the centre of mass positions and residue separations for the ΔL1-L3, ΔL1-P15, ΔY96 and ΔD89-Y96 mutants. All the mutants were studied at the air-water interface, whereas in the oil-water system we only considered the ΔL1-P15 mutant. For every mutant and type of interface we ran three independent simulations for 200 ns, a time that greatly exceeds the timescale of adsorption observed for the wild-type protein.

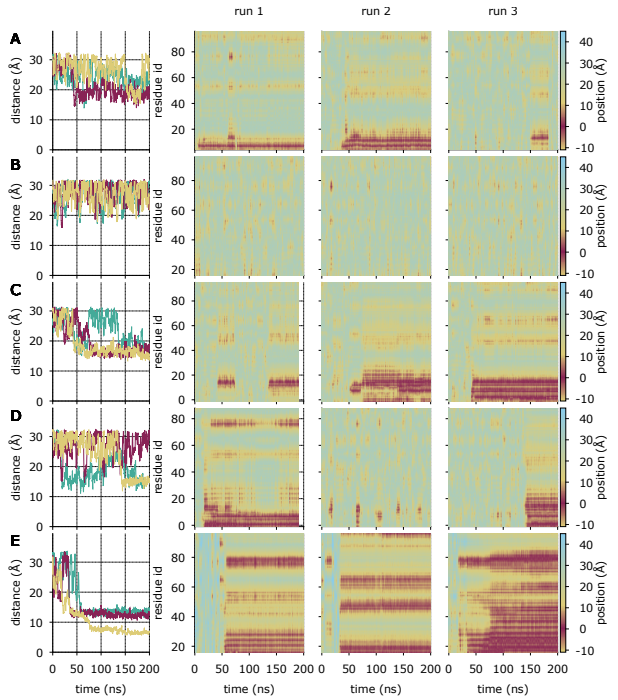

Fig. 6 Distance of the centre of mass from the interface and residue-interface separations for Rsn-2 mutants: (A) ΔL1-L3, (B) ΔL1-P15, (C) ΔY96 and (D) ΔD89-Y96 at the air-water interface, and (E) ΔL1-P15 at the oil-water interface. For the distance, cyan, red, and yellow lines denote first, second, and third runs respectively.

We will first consider the adsorption at the air-water interface. Deletion of most of the N-terminus (ΔL1-P15) is found to suppress adsorption completely over the length of the simulation. Even a smaller deletion (ΔL1-L3) is found to decrease the affinity of the protein for the interface. While protein adsorption is found in runs 1 and 2, the attachment to the interface occurs largely through the remaining flexible tail. The centre of mass separation once the protein has adsorbed is thus typically larger than for the wild type protein. In run 3 the protein adsorbs at 150 ns, but the contact is only temporary and the protein diffuses back into bulk water after 30 ns.

When the highly charged C-terminus is mostly removed (ΔD89-Y96), adsorption to the interface is not observed in one of the simulation runs. This result suggests that the more hydrophilic C-terminus aids adsorption, possibly by ensuring that the protein can achieve a favourable orientation to attach securely to the interface. Deleting only the final residue does not seem to have a large effect on the adsorption; the protein is able to adsorb to the interface within 200 ns in all three runs.

At the oil-water interface even the mutant expected to have the largest effect on the surface activity (ΔL1-P15) adsorbs in all three simulation runs. This is due to the fact that there are still many exposed hydrophobic residues in the protein (e.g. V16, I17, L20, F21, V78, V79), and that these residues interact with the cyclohexane more favourably than with the air phase, because of their higher partitioning free energies[31] (in the next section we will give a more detailed description of the energetics of the adsorption). Furthermore, in the third simulation run we observe the same unhinging conformational change found in the wild type protein. In this simulation, the unfolding is preceded by the formation of hydrophobic contacts between the cyclohexane molecules and residues I17, L20, F21, L25, F29, V78 and V79, similar to the behaviour observed in the wild-type protein. The observation that the unhinging transition at the interface proceeds via the same pathway of hydrophobic interactions suggests that the precise orientation of the protein at the interface is fundamental for this transition.

Figure 7 shows that the adsorption does not affect the secondary structure of the mutants, and even the unfolding of the hydrophobic core causes only a minor decrease in the number of α-helical residues. Again this suggests that Rsn-2 has specifically evolved to maintain its secondary structure at the interface.

3.3 Clustering of protein orientations

In order to further characterise the adsorption of Rsn-2, we performed a clustering analysis (see methods section) of the configurations observed after adsorption for the interface types and Rsn-2 mutants where the protein is able to adsorb in all three simulation runs: wt and ΔY96 Rsn-2 at the air/water interface, and wt and ΔL1-P15 Rsn-2 at the oil/water interface. As suggested from visual inspection of the simulation snapshots, the clustering confirms that the protein can interact with the interface in a limited number of orientations (Fig. 8). When it is folded, the protein adsorbs either parallel or perpendicular to the interface. The former is facilitated by the hydrophobic residues V78 and P79, whereas the latter is stabilised by residue L47. The third cluster corresponds to the unhinged and partially unfolded conformation of Rsn-2, where all the hydrophobic residues on the β-sheet and on the α-helix are able to interact with the interface. It is interesting to note that the two possible parallel and perpendicular orientations of the folded protein can be found in all types of interfaces and Rsn-2 variants considered. In most simulation runs, the protein maintains the same orientation, signalling the presence of a free energy barrier to changing orientation.

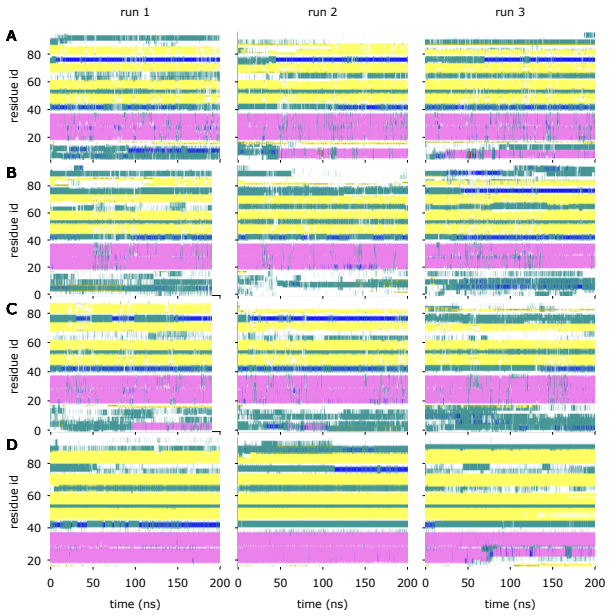

Fig. 7 Secondary structure per residue as a function of time for the Rsn-2 mutants: (A) ΔL1-L3, (B) ΔY96 and (C) ΔD89-Y96 at the air-water interface, and (D) ΔL1-P15 at the oil-water interface. We excluded ΔL1-P15 Rsn-2 at an air-water interface from the figure, since it does not adsorb in any of the considered runs and it is stable in the bulk.

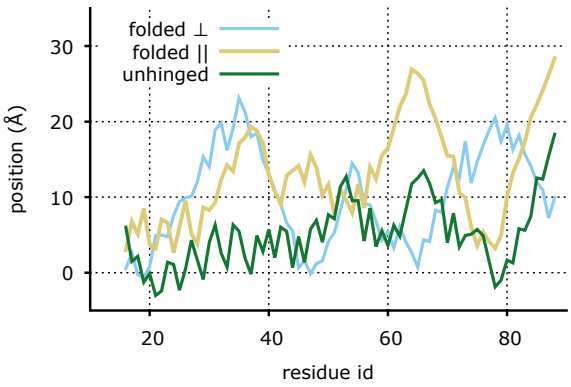

Fig. 8 We plot the Rsn-2 residue-interface separations (for residue ids 16 to 88) corresponding to the cluster centres of the three main clusters found in our analysis. Each cluster represents a different folding and orientation adopted by the protein adsorbed at the interface: in the cyan cluster the protein is folded and perpendicular to the interface, in the yellow one it is folded and parallel, and in the green one Rsn-2 is unhinged.

3.4 What drives interfacial adsorption?

More insight into the driving force for interfacial adsorption can be found by considering separate contributions to the adsorption free energy, specifically from the partitioning of the hydrophobic residues into the apolar (air or cyclohexane) phase and the change in interfacial free energy caused by the adsorption of the protein at the interface. The change in free energy of the system due to interfacial adsorption can be estimated as[36,37]:

$$\Delta F_{inter} = \Delta F_{partition} + \Delta F_{inter} = \sum_{i,\text{apolar}} e_i - \gamma \Delta A. \qquad (3)$$

where e_i is the free energy of transfer of residue i from the water to the apolar phase, the sum running over the residues located in the apolar phase, γ is the air-water or cyclohexane-water interfacial tension, and ΔA is the area of interface occupied by the protein. The first term, the partitioning energy $\Delta F_{\text{partition}}$, quantifies the gain (or loss) in energy when a residue is transferred from the initial water phase to the air or oil phase[36], and its optimisation tends to move hydrophobic residues into the apolar phase, and to keep hydrophilic residues into the water phase. The second term, the interfacial energy ΔF_{inter}, represents the cost associated with the presence of an interface between two immiscible fluids[37]. As the protein occupies the interface, it lowers the energy of the system by reducing the interfacial area between the two fluids.

Presented in Table 1 are the values of $F_{partition}$, ΔA, and F_{inter} for Rsn-2 at both air-water and oil-water interfaces. The interfacial free energies were calculated using simulation values of the interfacial tension; for the air-water interface $\gamma = 52.3$ mN/m[33] and for the water-cyclohexane interface $\gamma = 42.7$ mN/m (see methods). At the air-water interface the change in interfacial free energy upon adsorption is significantly larger than the partition free energy. On the other hand, at the oil-water interface partitioning and interfacial contributions to the adsorption are comparable, especially for the wild type protein. This is explained by the fact that hydrophobic residues gain more energy from being transferred to the cyclohexane phase than to the air phase, for instance the partition free energy of leucine is 4.92 kcal/mol from cyclohexane to water and 2.28 kcal/mol from air to water[31]; some residues even switch from being hydrophobic to hydrophilic, e.g. the partition free energy of phenylalanine is 2.98 kcal/mol from cyclohexane to water and -0.76 kcal/mol from air to water[31]. This observation helps to explain firstly why N-terminal deletions are sufficient to affect the adsorption at the air-water interface but not at the oil-water interface, and secondly why we observe the unfolding and exposure of the protein core to the hydrophobic phase only at the oil-water interface.

Comparison of the surface areas shows that the protein occupies a significantly larger area at the water-cyclohexane interface than the air-water interface. This again is a consequence of the stronger partitioning of the hydrophobic residues out of the water. For the runs where unhinging is observed (wt run1 and ΔL1-P15 run 3) the interfacial area is approximately two to three times larger than that measured for simulations where the protein remains in a closed conformation. The areas from the simulations for the unhinged proteins estimated using this method are smaller than those found from neutron reflectivity[2] (~ 1412 Å^2).

Finally, we analysed the unhinging of the hydrophobic core at the oil-water interface (run 1 for wild type Rsn-2 and run 3 for the ΔL1-P15 mutant) by looking at the number of contacts (see methods) formed between the cyclohexane molecules and the hydrophobic residues in the core of the protein (core contacts) and the contacts between the hydrophobic residues on the β-sheet and those on the α-helix of the protein ($\alpha - \beta$ contacts). Figure 9 shows that the adsorption of the protein corresponds to the formation of the first contacts with the cyclohexane molecules,

	Run	$\Delta F_{partition}$ (kcal mol^{-1})	A (Å^2)	$\gamma_{sim}A$ (kcal mol^{-1})
Air-water interface				
Wild type	1	-0.80±0.09	142.4±1.3	10.72±0.10
	2	-1.27±0.05	70.6±0.8	5.31±0.06
	3	-1.38±0.09	125.9±1.2	9.48±0.09
ΔL1-L3	1	-0.317±0.022	19.9±0.4	1.50±0.05
	2	-0.48±0.04	46.2±0.6	3.48±0.11
	3	0.021±0.013	2.6±0.2	0.192±0.018
ΔD89-Y96	1	-2.41±0.04	48.4±0.4	3.6±0.11
	2	-0.009±0.003	0.93±0.13	0.070±0.010
	3	-2.58±0.06	124.4±1.2	9.4±0.3
ΔY96	1	-0.70±0.03	26.9±0.3	2.02±0.06
	2	-2.58±0.07	100.8±1.0	7.6±0.2
	3	-1.46±0.08	124.5±1.2	9.4±0.3
Cyclohexane-water interface				
Wild type	1	-20.65±0.54	524.2±4.3	32.22±0.26
	2	-27.28±0.34	278.0±1.7	17.09±0.10
	3	-5.20±0.35	238.9±2.5	14.68±0.15
ΔL1-P15	1	0.43±0.15	82.3±2.5	5.06±0.15
	2	0.66±0.17	128.2±3.2	7.88±0.20
	3	-5.69±0.31	442.0±3.6	27.16±0.22

Table 1 Average partition free energies, occupied surface areas, and interfacial free energies for Rsn-2 at air-water interface. Average values calculated over final 20 ns of simulations, uncertainties estimated from standard error

while the $\alpha-\beta$ contacts remain stable at their initial value. Then, around 180 ns for wild type Rsn-2 and around 65 ns for the mutant, we observe a further increase in the number of oil contacts and a simultaneous decrease in the number of contacts between the α-helix and the β-sheet. These plots confirm what was already suggested from the observation of the simulation trajectory: the hydrophobic interface between the helix and the β-sheet is directly replaced with a new interface with the oil without having to first expose the residues to the unfavourable water solvent. This process is not expected to be significantly favourable nor unfavourable; therefore the unhinging appears mainly driven by the interfacial energy gained from the large increase in the area occupied by the protein at the interface (see Table 1).

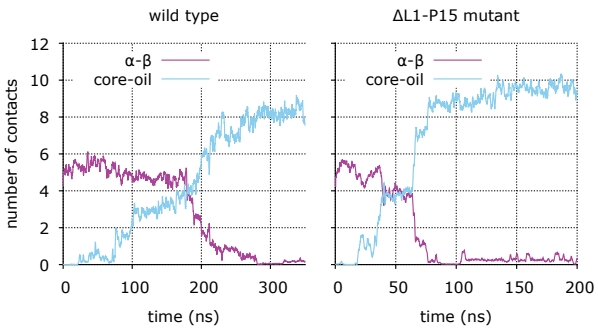

Fig. 9 Number of hydrophobic contacts within the core of the protein, i.e. formed between the β-sheet and the α-helix, (purple), and between the hydrophobbhobic residues in the core and the cyclohexane molecules (cyan). The plots correspond to the two events where the unhinging transition is observed: (left) run 1 for wild type Rsn-2 and (right) run 3 for ΔL1-P15 Rsn-2.

4 Discussion and Conclusions

Protein biosurfactants comprise a diverse group of proteins that exhibit a number of structural features that allow them to adsorb onto liquid interfaces while remaining biocompatible and avoiding aggregation in solution. Because of this biosurfactants exhibit a range of structures beyond the simple hydrophobic tail-hydrophilic head structure of synthetic surfactants and often undergo specific conformational changes upon interfacial adsorption. One example of this is the protein Rsn-2, whose structure unhinges upon adsorption at interfaces.

Using molecular dynamics simulations we have investigated the adsorption of Rsn-2 onto the air-water and cyclohexane-water interfaces, in order to investigate the initial stages of adsorption and in the case of the cyclohexane-water interface the unhinging transition. In both cases the protein adsorbs through the flexible, hydrophobic N-terminus, indicating the importance of this region for the interfacial attachment, similar to the fly-casting mechanism found in disordered proteins[35]. Permanent attachment, however, involves additional hydrophobic contacts between the protein and interface. For the cyclohexane-water interface unhinging of the protein was observed which was absent for the air-water interface. This difference in behaviour is likely to arise due to the more favourable partition energies of the hydrophobic residues into cyclohexane compared to air and the ability of the oil molecules to penetrate the protein core, destabilising the closed conformation. For the air-water interface there are no hydrophobic molecules that can act to destabilise the closed conformation so the opening will occur over a longer timescale (longer than the simulation timescales).

The effect of the N- and C-termini on interfacial adsorption was explained through simulations of Rsn-2 mutants with por-

tions of these regions deleted. Removal of the entire N-terminal tail ($\Delta L1 - P15$) was found to prevent adsorption at the air-water interface across the entire length of the simulation. Smaller deletions also affected adsorption, with one of the $\Delta L1 - L3$ simulation runs failing to adsorb. C-terminal deletions have a weaker effect on the adsorption, with the ΔY96 mutant showing similar behaviour to the wild type protein. This is consistent with experimental dynamic surface tension measurements, where mutants with the $\Delta L1 - P15$ deletion are found to both have higher limiting surface tension and slower decreases in the surface tension, indicating that the speed and extent of interfacial adsorption is lower for these. The ΔL1-L3 mutant also exhibits a higher limiting surface tension compared to the wild type protein, although the effect is smaller than for ΔL1-P15 mutant, which again is consistent with the simulation studies. The fact that neither mutation at the C-terminus has any effect upon the final surfactant activity of Rsn-2 supports the idea that the C-terminus remains within the solvent and that the solvent exposed face is sufficiently hydrophilic without the poly-aspartic acid motif.

Compared to the air-water interface protein, deletions have a smaller effect at the cyclohexane-water interface. The $\Delta L1 - P15$ mutant still adsorbs at the cyclohexane-water interface, as the partition free energy for the remaining hydrophobic residues is more favourable at this interface.

The difference between adsorption at air-water and cyclohexane-water interfaces can be better understood through considering the individual partitioning and interfacial contributions to the free energy of the system; for the oil-water interface these are comparable to each other whereas the interfacial energy is more important for the air-water interface, due to the combination of higher surface tension for the air-water interface and lower air-water transfer free energies. While it is not directly involved in the adsorption, mutations involving C-terminal deletions also affect the adsorption at the air-water interface, potentially as this region helps orient the protein to enhance the attachment probability.

By determining the residues involved in the adsorption of Rsn-2 at liquid interfaces and examining the effect of protein mutations on this we have provided a detailed characterisation of its interfacial adsorption. This will help inform new applications of proteins in areas including food products[38] and drug delivery[39] and in the design of synthetic molecules that exhibit similar conformational changes at liquid interfaces[40].

Acknowledgements

Computational resources for this work were provided by ARCHER granted via the UK High-End Computing Consortium for Biomolecular Simulation (HECBioSim, EPSRC grant no. EP/L000253/1), the ARCHIE-WeSt High Performance Computing facility, University of Strathclyde (EPSRC grant no. EP/K000586/1), and SFI/HEA Irish Centre for High-End Computing (ICHEC). The research was additionally supported by BB/L006979 and EP/J007404. GBB has been funded by the Principals Career Development Scholarship of the University of Edinburgh. SJV was supported by a studentship from the Biotechnology and Biological Sciences Research Council.

Appendix - Simulations at air-water interface using Amber99sb-ildn force field

Shown in Fig. 10 are the protein centre-of-mass-interface and residue-interface separations from simulations of Rsn-2 at an air-water interface using the Amber99sb-ildn force field. This shows similar behaviour to simulations using the Charmm-27 force field (Fig. 1). The protein makes transient contact with the interface after about 10 ns. It then contacts the interface at approximately 26 ns, remaining attached for the remainder of the simulation. As for the Charmm-27 simulations this attachment again is mediated by the N-terminus. Further residues (around residues 45 and 60) are in contact with the interface, with these being further from the interface during the initial, transient contact.

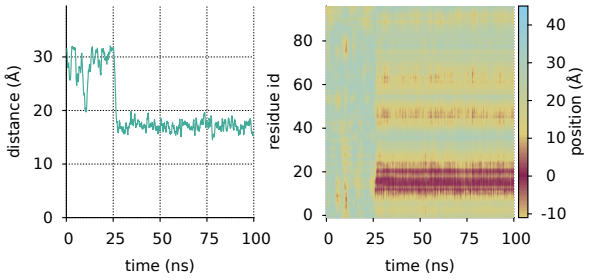

Fig. 10 Distance of the protein centre of mass from an air-water interface as a function of time and residue-interface separations from simulations using Amber99sb-ildn force field.

As for the Charmm-27 simulations at an air-water interface, as well as simulations at a water-cyclohexane interface the secondary structure remains largely unchanged during adsorption at an air-water interface (Fig. 11).

This demonstrates that the behaviour of the protein is similar using both Charmm-27 and Amber99sb-ildn force fields, suggesting that the observed behaviour is not dependent on the force field used.

References

1 J. Zasadzinski, J. Ding, H. Warriner, F. Bringezu and A. J. Waring, *Curr. Opin. Colloid Interface Sci.*, 2001, **6**, 506–513.

2 C. D. Mackenzie, B. O. Smith, A. Meister, A. Blume, X. Zhao, J. R. Lu, M. W. Kennedy and A. Cooper, *Biophys. J.*, 2009, **96**, 4984–92.

3 M. B. Linder, *Curr. Opin. Colloid Interface Sci.*, 2009, **14**, 356–363.

4 A. Cooper and M. W. Kennedy, *Biophys. Chem.*, 2010, **151**, 96–104.

5 M. Schor, J. L. Reid, C. E. MacPhee and N. R. Stanley-Wall, *Trends Biochem. Sci.*, 2016.

6 K. M. Bromley, R. J. Morris, L. Hobley, G. Brandani, R. M. C. Gillespie, M. McCluskey, U. Zachariae, D. Marenduzzo, N. R. Stanley-Wall and C. E. MacPhee, *Proc. Natl. Acad. Sci. USA*, 2015, 5419–5424.

7 G. B. Brandani, M. Schor, R. Morris, N. Stanley-Wall, C. E.

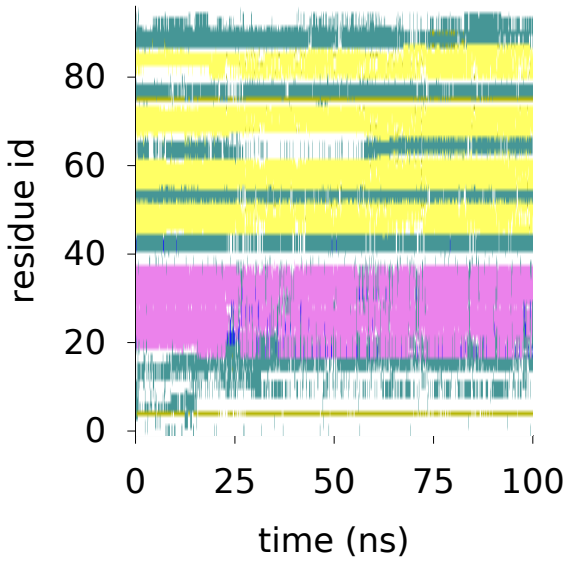

Fig. 11 Protein secondary structure as function of time at an air-water interface from simulations using Amber99sb-ildn force field.

MacPhee, D. Marenduzzo and U. Zachariae, *Langmuir*, 2015, **31**, 11558–63.

8 D. G. Dalgleish, *Soft Matter*, 2011, **7**, 2265–2272.

9 A. Cooper, M. W. Kennedy, R. I. Fleming, E. H. Wilson, H. Videler, D. L. Wokosin, T.-J. Su, R. J. Green and J. R. Lu, *Biophys. J.*, 2005, **88**, 2114–25.

10 R. J. Morris, G. B. Brandani, V. Desai, B. O. Smith, M. Schor and C. E. MacPhee, *Biophys. J.*, 2016.

11 D. L. Cheung, *Langmuir*, 2012, **28**, 8730–8736.

12 A. De Simone, C. Kitchen, A. H. Kwan, M. Sunde, C. M. Dobson and D. Frenkel, *Proc. Natl. Acad. Sci. USA*, 2012, **109**, 6951–6.

13 S. R. Euston, *Food Hydrocoll.*, 2013, **42**, 66–74.

14 D. Zare, K. M. McGrath and J. R. Allison, *Biomacromolecules*, 2015, **16**, 1855–61.

15 D. Zare, J. R. Allison and K. M. McGrath, *Biomacromolecules*, 2016, **17**, 1572–1581.

16 S. R. Euston, P. Hughes, M. A. Naser and R. E. Westacott, *Biomacromolecules*, 2008, **9**, 1443–53.

17 M. Arooj, N. S. Gandhi, C. A. Kreck, D. W. Arrigan and R. L. Mancera, *J. Phys. Chem. B*, 2016, **120**, 3100–3112.

18 D. L. Cheung, *Langmuir*, 2016, **32**, 4405–4414.

19 B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl, *J. Chem. Theory Comput.*, 2008, **4**, 435–447.

20 A. D. MacKerell, D. Bashford, M. Bellott, R. Dunbrack, J. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. a. Ha *et al.*, *J. Phys. Chem. B*, 1998, **102**, 3586–3616.

21 A. D. Mackerell, M. Feig and C. L. Brooks, *J. Comput. Chem.*, 2004, **25**, 1400–15.

22 K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw, *Proteins*, 2010, **78**, 1950–1958.

23 A. Cordomi, G. Caltabiano and L. Pardo, *J. Chem. Theory Comput.*, 2012, **8**, 948–958.

24 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys*, 1983, **79**, 926.

25 G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys*, 2007, **126**, 14101.

26 M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182–7190.

27 B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *J. Comput. Chem.*, 1997, **18**, 1463–1472.

28 S. Miyamoto and P. A. Kollman, *J. Comput. Chem.*, 1992, **13**, 952–962.

29 T. Darden, D. York and L. Pedersen, *J. Chem. Phys*, 1993, **98**, 10089.

30 D. Frishman and P. Argos, *Proteins*, 1995, **23**, 566–579.

31 A. Radzicka and R. Wolfenden, *Biochemistry*, 1988, **27**, 1664–1670.

32 R. L. C. Vink, J. Horbach and K. Binder, *J. Chem. Phys.*, 2005, **122**, 134905.

33 C. Vega and E. de Miguel, *J. Chem. Phys.*, 2007, **126**, 154707.

34 A. Rodriguez and A. Laio, *Science*, 2014, **344**, 1492–1496.

35 Y. Levy, J. N. Onuchic and P. G. Wolynes, *J. Am. Chem. Soc.*, 2007, **129**, 738–9.

36 D. Eisenberg, R. M. Weiss, T. C. Terwilliger and W. Wilcox, Faraday Symp. Chem. Soc., 1982, pp. 109–120.

37 R. Aveyard, B. P. Binks and J. H. Clint, *Adv. Colloid Interface Sci.*, 2003, **100-102**, 503–546.

38 E. Dickinson, *Soft Matter*, 2006, **2**, 642.

39 B. J. Zeng, Y. P. Chuan, B. O'Sullivan, I. CaminschiI, M. H. Lahoud, R. Thomas and A. P. J. Middelberg, *Small*, 2013, **9**, 3736–3742.

40 D. Ishikawa, T. Mori, Y. Yonamine, W. Nakanishi, D. L. Cheung, J. P. Hill and K. Ariga, *Angew. Chemie Int. Ed.*, 2015, **54**, 8988–8991.